Appendix

Appendix A1 Study characteristics: Sáenz, Fuchs, & Fuchs, 2005 (randomized controlled trial)

Characteristic	Description
Study citation	Sáenz, L. M., Fuchs, L. S., & Fuchs, D. (2005). Peer-Assisted Learning Strategies for English language learners with learning disabilities. *Exceptional Children,* 71, 231–247.
Participants	A total of 132 native Spanish-speaking students in grades 3–6 participated in the study. At least two students in each participating classroom had to have a learning disability, and all students had to be English language learners. Twelve classrooms were included in the study, and, prior to random assignment of classrooms to conditions, they were stratified by grade level and campus. A total of 119 students were included in posttest analyses (two students with a learning disability, three low-achieving students, three average-achieving students, and three high-achieving students in each of the 12 classrooms).[1] All teachers taught reading (only) in transitional bilingual educational classrooms and had at least a bachelor's degree. Additionally, they were all certified in ELL or bilingual instruction. Each teacher had two classes, although only one was included in the study.
Setting	Students attended schools in South Texas. All students were enrolled in bilingual education classrooms.
Intervention	*Peer-Assisted Learning Strategies* was implemented three times a week for 15 weeks. Each peer-assisted learning session lasted for 25–35 minutes and occurred during regular reading instruction periods. Teachers ranked students by their reading achievement (high versus low) and paired a higher-achieving student with a lower-achieving student. Students were assigned a new partner about once a month. During each lesson, students took turns acting as the tutor and tutee as they participated in three reading activities: partner reading with story retell, paragraph shrinking, and prediction relay. Pairs earned points for correct or accurate responses during activities. Students in *Peer-Assisted Learning Strategies* classrooms received one-on-one instruction for 26 percent of activities. During student training, research assistants provided daily technical assistance and then weekly assistance after the training.
Comparison	Teachers in the comparison group provided the district's regular curriculum for reading instruction, which consisted of little one-on-one peer instruction and was mainly teacher-led. Lesson plans for both the intervention and comparison classrooms were reviewed twice during the study to assess the type of instruction provided. In comparison classrooms, 13 percent of activities were conducted on a one-on-one basis.
Primary outcomes and measurement	Reading achievement was assessed using the Comprehensive Reading Assessment Battery (CRAB) (see Appendix A2 for more detailed descriptions of outcome measures).
Teacher training	Teachers were taught how to train their students on *Peer-Assisted Learning Strategies* during a full-day workshop. They were also provided with an overview of procedures associated with the intervention, practiced intervention activities, and were given a *Peer-Assisted Learning Strategies* manual. The manual included scripted lessons for teachers to use while training students on the intervention.

1. Overall sample attrition equaled 10 percent at the student level; the report indicates that 13 students left the study either because of relocation to another school in the district, or they went elsewhere in the United States because of seasonal employment. No teachers left the study.

Appendix A2 Outcome measures in the reading achievement domain

Characteristic	Description
Comprehensive Reading Assessment Battery (CRAB)	The Comprehensive Reading Assessment Battery (CRAB) includes four 400-word folktales. Students have three minutes to read the first folktale aloud and then answer ten comprehension questions. For a second folktale, students have two minutes to complete a cloze or maze task, three minutes to read the story aloud, and then answer ten comprehension questions.
CRAB: Words Correct Subscale	For the words correct subscale, reading fluency and accuracy are assessed. Scores on this measure were based on the number of words read correctly in three minutes (measure of reading fluency). According to the study authors, the measure has demonstrated reliability and validity.
CRAB: Comprehension Questions Correct Subscale	The comprehension questions correct subscale assesses reading comprehension. Scores were based on the number of correct answers to comprehension questions. According to the study authors, the measure has demonstrated reliability and validity.
CRAB: Maze Choices Correct Subscale	The maze choices correct subscale assesses silent reading accuracy and fluency. The maze task requires students to read a passage that consists of the first sentence intact, followed by every seventh word replaced with a 3-item multiple-choice format. One choice is a semantically (grammatically and contextually) correct replacement for the missing word. Scores on this measure were based on the number of correct maze choices made in two minutes. According to the study authors, the measure has demonstrated reliability and validity.

Appendix A3 Summary of study findings included in the rating for the reading achievement domain[1]

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *PALS* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*PALS* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Sáenz, Fuchs, & Fuchs, 2005 (randomized controlled trial)[7]								
CRAB: Words Correct	Grades 3–6	119	26.06 (93.35)	10.83 (103.03)	15.23	0.15	ns	+6
CRAB: Comprehension Questions Correct	Grades 3–6	119	0.95 (2.41)	–0.42 (1.73)	1.37	0.64	ns	+24
CRAB: Maze Choices Correct	Grades 3–6	119	1.59 (4.33)	0.93 (4.02)	0.66	0.16	ns	+6
Domain average[8] for reading achievement						0.32	ns	+12

ns = not statistically significant

1. This appendix reports the findings considered for the effectiveness rating and the average improvement indices. Subgroup findings, which were the main focus of the Sáenz et al. (2005) study, are not included here, but are reported in Appendix A4.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are; a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. The means and standard deviations are aggregates of the achievement categories originally reported in the Sáenz et al. (2005) study.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. For the purposes of this report, pretest-posttest difference scores are presented and used to calculate estimates of effects.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations. Student-level effect sizes were calculated for the purposes of this review. Student-level means and standard deviations by student type (such as learning disabled) and across student type for each measure were obtained from the primary study author.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Sáenz et al. (2005), corrections for clustering and multiple comparisons were needed, so the significance levels differ from those reported in the original study.
8. This row provides the study average, which in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4 Summary of Grades 3–6 subgroup findings for the reading achievement domain[1]

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *PALS* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*PALS* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Sáenz, Fuchs, & Fuchs, 2005 (randomized controlled trial)[7]								
CRAB: Words Correct	Learning disabled	20	26.65 (81.17)	–6.35 (92.36)	33.00	0.36	ns	+14
CRAB: Comprehension Questions Correct	Learning disabled	20	1.15 (1.43)	–0.15 (1.31)	1.30	0.91	ns	+32
CRAB: Maze Choices Correct	Learning disabled	20	0.90 (3.16)	–0.60 (2.98)	1.50	0.47	ns	+18
CRAB: Words Correct	Low achieving	33	22.00 (80.97)	14.03 (78.85)	7.97	0.10	ns	+4
CRAB: Comprehension Questions Correct	Low achieving	33	0.77 (2.60)	–0.14 (1.55)	0.91	0.42	ns	+16
CRAB: Maze Choices Correct	Low achieving	33	1.40 (3.75)	1.22 (3.95)	0.18	0.05	ns	+2
CRAB: Words Correct	Average achieving	35	12.97 (66.98)	8.44 (75.32)	4.53	0.06	ns	+2
CRAB: Comprehension Questions Correct	Average achieving	35	0.74 (1.76)	–0.42 (1.37)	1.16	0.72	ns	+26
CRAB: Maze Choices	Average achieving	35	1.94 (3.66)	2.00 (3.18)	–0.06	0.02	ns	+1
CRAB: Words Correct	High achieving	31	42.38 (68.00)	22.07 (91.29)	20.31	0.25	ns	+10
CRAB: Comprehension Questions Correct	High achieving	31	1.21 (1.71)	–0.96 (1.78)	2.17	1.21	Statistically significant	+39
CRAB: Maze Choices Correct	High achieving	31	1.82 (4.26)	0.29 (4.20)	1.53	0.35	ns	+14

ns = not statistically significant

1. This appendix presents subgroup findings for measures that fall in the reading achievement domain. Total group scores were used for rating purposes and are presented in Appendix A3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.

(continued)

Appendix A4 Summary of Grades 3–6 subgroup findings for the reading achievement domain *(continued)*

3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. For the purposes of this report, pretest-posttest difference scores are presented and used to calculate estimates of effects.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations. Student-level effect sizes were calculated for the purposes of this review. Student-level means and standard deviations by student type (learning disabled) and across student type for each measure were obtained from the primary study author. Means and standard deviations presented in Appendix A3 are aggregates of the means and standard deviations presented in this appendix.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Sáenz et al. (2005), a correction for clustering was needed, so the significance levels differ from those reported in the original study.

Appendix A5 *Peer-Assisted Learning Strategies* rating for the reading achievement domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of reading achievement, the WWC rated *Peer-Assisted Learning Strategies* as potentially positive. It did not meet the criteria for positive effects because it only had one study. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered because *Peer-Assisted Learning Strategies* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Met. One study showed substantively important positive findings.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Met. No study of *Peer-Assisted Learning Strategies* showed a statistically significant or substantively important negative effect or indeterminate effect.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. Only one study was reviewed.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. The one study reviewed did not show statistically significant or substantively important negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Schools	Students	Extent of evidence[1]
Reading achievement	1	nr	132	Small
Mathematics achievement	0	0	0	na
English language development	0	0	0	na

na = not applicable/not studied
nr = not reported

1. A rating of "moderate to large" requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is "small."